Naval Health Research Center

BLIND SEPARATION OF EVENT-RELATED BRAIN RESPONSES INTO INDEPENDENT COMPONENTS

S. Makeig
T-P Jung
A. J. Bell
D. Ghahremani
T. J. Sejnowski

Report No. 96-10

NAVAL HEALTH RESEARCH CENTER
P.O. BOX 85122
SAN DIEGO, CALIFORNIA 92186-5122

NAVAL MEDICAL RESEARCH AND DEVELOPMENT COMMAND
BETHESDA, MARYLAND

Blind Separation of Event-related Brain Responses into Independent Components[1]

Scott Makeig†○, Tzyy-Ping Jung†‡,
Anthony J. Bell‡, Dara Ghahremani‡,
Terrence J. Sejnowski‡§

May 21, 1996

†Naval Health Research Center
P.O. Box 85122
San Diego, CA 92186-5122

‡Howard Hughes Medical Institute
Computational Neurobiology Laboratory
The Salk Institute, P. O. Box 85800
San Diego, CA 92186-5800

○ Department of Neurosciences
School of Medicine
University of California San Diego
La Jolla, CA 92093

§Department of Biology
University of California San Diego
La Jolla, CA 92093

REPORT NO. 96-10

[1]This report was supported in part by the Navy Medical Research and Development Command and the Office of Naval Research, Department of the Navy under work unit ONR Reimb-6429. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, Department of Defense, or the U.S. Government. Approved for public release, distribution unlimited.

Functional imaging of brain activity based on changes in blood flow does not supply information about the relative timing of brief bursts of neural activity in different brain areas[1, 2]. Multichannel electric or magnetic recordings from the scalp provide high temporal resolution, but are not easily decomposed into the separate activities of multiple brain networks. We report here a method for the blind separation of event-related brain responses into spatially stationary and temporally independent subcomponents using an Independent Component Analysis algorithm[3]. Applied to electroencephalographic responses from an auditory detection task[4], each of the most active identified sources accounted for all or part of a previously identified response component. This spatiotemporal decomposition was robust to changes in sensors and input data length, and was stable within subjects. The method can be used to assess the timing, strength, and stability of event-related activity in brain networks during cognitive tasks, regardless of source location.

Electroencephalographic (EEG) and magnetoencephalographic (MEG) signals recorded from scalp sensors register the integrated activity arising from a large number of local neural processes whose temporal coherence is highly variable[5]. Event-related potentials (ERPs) elicited by sensory stimuli or time locked to experimental events represent a sum of subcomponents originating in different brain areas[6]. These subcomponents may overlap temporally and spatially, making their separation problematic and their localization ambiguous[7]. Here, we propose a method for determining these subcomponents under the assumption that EEG and MEG signals, including evoked responses, arise from mutually independent activities of neurons in a small number of spatially stationary brain networks.

Previous decompositions of evoked responses using principal component analysis used only second-order correlational information and assumed each source had the same time course of activation in every experimental condition[8]. The algorithm we use is based on an 'infomax' neural network[3, 9, 10]. It finds, by stochastic gradient ascent, a matrix, $\mathbf{W}$, which maximizes the entropy[11], $H(\mathbf{y})$, of an ensemble of 'sphered' input vectors $\{\mathbf{x}_s\}$, linearly transformed and sigmoidally compressed ($\mathbf{u} = \mathbf{W}\mathbf{x}_s$, $\mathbf{y} = g(\mathbf{u})$). The 'unmixing' matrix $\mathbf{W}$ performs source separation, while the sigmoidal nonlinearity $g()$ provides necessary higher-order statistical information. Initial sphering of the zero-mean input data[12] ($\mathbf{x}_s = \mathbf{P}\mathbf{x}$, where $\mathbf{P} = 2\langle \mathbf{x}\mathbf{x}^T \rangle^{-\frac{1}{2}}$) speeds convergence. $\mathbf{W}$ is then initialized to the identity matrix ($\mathbf{I}$) and iteratively adjusted using small batches of data vectors drawn randomly from $\{\mathbf{x}_s\}$ without substitution, according to:

$$\Delta\mathbf{W} = \epsilon \frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} \mathbf{W}^T\mathbf{W} = \epsilon(\mathbf{I} + \hat{\mathbf{y}}\mathbf{u}^T)\mathbf{W}, \qquad (1)$$

where ϵ is the learning rate and vector $\hat{\mathbf{y}}$ has elements $\hat{y}_i = (\partial/\partial u_i)\ln(\partial y_i/\partial u_i)$ The $(\mathbf{W}^T\mathbf{W})$ 'natural gradient' term in the update equation[13] avoids matrix inversions

and speeds convergence. We use the logistic nonlinearity, $g(u_i) = (1 + \exp(-u_i))^{-1}$, for which $\hat{y}_i = 1 - 2y_i$. Further details and references about the algorithm appear in [3, 13, 14], other related approaches and background material in [10, 15, 16, 17, 18].

When ICA is trained on EEG data (Fig. 1), the rows of the resultant matrix ($\mathbf{WP}$) are linear spatial filters which, applied to the input data, produce source activity waveforms ($\mathbf{WPx}$). The columns of the inverse weight matrix $(\mathbf{WP})^{-1}$ represent the projection weights from the ICA sources to the sensor array. ICA sources may be spatially distributed, so they need not correspond to single peaks in functional imaging measurements. The minimum number of time points needed for the method appears to be several times the number of recording channels, which in turn must be larger than the number of sources to be separated.

We first applied the ICA algorithm to 300-s EEG epochs to demonstrate that the algorithm can segregate brain from muscle activity and can track psychophysiological state changes[19]. To test the performance of the ICA algorithm in separating multiple strong and weak brain sources, we then performed numerical simulations using a three-shell spherical head model[20]. Results suggested that the algorithm can successfully separate the independent activities of stronger brain sources from those of weaker signal and noise sources, even when they overlap in time and space[21].

Next, we applied the ICA algorithm to 1-s ERPs time locked to the detected and undetected noise-burst targets (Fig. 2a) presented in an experiment in which a subject was asked to detect weak noise-burst targets continuously during a half-hour session[4]. Results of the analysis consisted of 15 ICA sources (Fig. 2b), including one source (U1) accounting for the positive 320-ms component of the response to undetected targets and a separate source (D1) accounting for the positive 320-ms detected-

target response component. Three more sources (U2-4) with overlapping activations accounted for the broad negativity peaking at 432 ms in the undetected-target response. Source D2 accounted for the positive detected-target response component peaking near the median motor response latency (406 ms). Three other sources (S1-3) together composed the auditory steady-state response (SSR)[22] produced by a 39 Hz click train delivered continuously during the experiment. The projected scalp distributions of all these sources were distinct (although not orthogonal) and summed to the scalp distributions of the corresponding response components (Fig. 3). We presume the four remaining sources mixed activity from multiple weak sources.

Both the activity waveforms and the scalp projections of the ICA sources with largest scalp projections proved robust to changes in initial weights, number and placement of electrodes, epoch lengths, and number of training conditions (Fig. 4). Nearly identical ICA sources were recovered from responses collected on different days from the same subject (Figs. 3a, 3b, 4a), and similar sources from different subjects in the same experiment (Fig. 4b). ICA accounted for the conventionally-defined ERP components of the data (major response peaks and the SSR) as sums of activity of one to three ICA sources. Although the algorithm used no temporal sequence information, several of the identified sources were active only during a brief interval (50-300 ms), suggesting that their activity may represent spatiotemporally distinct processing steps.

ICA can be applied conveniently to event-related electric or magnetic responses from any number of available channels and experimental conditions. It attempts to measure *what* independent activities compose its training data without directly specifying *where* these activities are located. The method appears promising for

measuring the effects of experimental variables on evoked-response subcomponents representing rapid stages of brain information processing, particularly when these overlap in time and spatial projection.

References and notes

1. Aine, C. J. *Crit. Rev. Neurobiol.* **9**, 229-309 (1995).

2. Snyder, A. Z., Abdullaev, Y. G., Posner, M. I. & Raichle, M. E. *Proc. Natl. Acad. Sci. USA* **92**, 1689-1693 (1995).

3. Bell, A.J. & Sejnowski, T.J. *Neural Computation* **7**, 1129-1159 (1995).

4. Makeig, S. & Inlow, M. *Electroencephalogr. clin. Neurophysiol.* **86**, 23-35 (1993)

5. Bullock, T. H., McClune, M. S., Achnimowicz, J. Z., Iragui-Madoz, V. J., Duckrow, R. B. & Spencer, S. S. *Proc. Natl. Acad. Sci. USA* **92**, 11568-11572 (1995).

6. Naatanen, R. & Picton, T. *Psychophysiology* **24**, 375-425 (1987).

7. Nunez, P.L. *Electric Fields of the Brain.* New York: Oxford (1981).

8. Chapman, R.M. & McCrary, J.W. *Brain Lang.* **27**, 288-301 (1995).

9. Linsker, R. *Neural Computation* **4**, 691-702 (1992).

10. Nadal, J-P. & Parga, N. *Network* **5**, 565-581 (1994).

11. Cover, T.M. & Thomas, J.A. *Elements of Information Theory*, John Wiley (1991).

12. Bell, A.J. & Sejnowski, T.J. *Network: Computation in Neural Systems*, **7**, 2 (1996).

13. Amari S., Cichocki, A. & Yang, H.H. *Advances in Neural Information Processing Systems* 8, MIT Press (1996).

14. Bell, A.J. & Sejnowski, T.J. *Proc. Intern. Symp. on Nonlinear Theory and Applications*, Las Vegas (1995).

15. Comon, P. *Signal Processing* **36**, 287-314 (1994).

16. Jutten, C. & Herault, J. *Signal Processing* **24**, 1-10 (1991).

17. Karhumen, J., Oja, E., Wang, L., Vigario, R. & Joutsenalo, J. *IEEE Trans. Neural Networks* (to appear).

18. Cardoso, J-F. & Laheld, B. *IEEE Trans. Signal Proc.* (to appear).

19. Makeig, S., Bell, A.J., Jung, T-P. & Sejnowski T.J. *Advances in Neural Information Processing Systems* **8**, MIT Press (1996).

20. Dale, A.M. & Sereno, M.I. *J. Cogn. Neurosci.* **5**, 162-176 (1993).

21. Ghahremani, D., Makeig, S., Jung, T-P., Bell, A.J. & Sejnowski, T.J. *Tech. Rep.*, Institute for Neural Computation, San Diego, CA. (1996).

22. Galambos, R., Makeig, S. & Talmachoff P. *Proc. Natl. Acad. Sci. USA* **78**, 2643-2647 (1981).

23. Pantev, C., Elbert, T., Makeig, S., Hampson, S., Eulitz, C. & Hoke, M. *Electroencephalogr. clin. Neurophysiol.* **88**, 389-396 (1993).

24. Perrin, F., Pernier, J., Bertrand, O., & Echallier, J.F. *Electroencephalogr. clin. Neurophysiol.* **72**, 184-187 (1989).

Acknowledgments

This report was supported in part by grants to S.M., T-P.J. and T.J.S. from the Office of Naval Research, and to T.J.S. and A.J.B. from the Howard Hughes Medical Institute. The authors acknowledge the contributions of F.S. Elliott and M. Inlow in collecting and processing the data, and thank A. Dale for supplying the head model and S. Hillyard, L. Anllo-Vento, and J. Hansen for discussions and plotting assistance.

Figure Legends

Figure 1 - Schematic overview of Independent Component Analysis (ICA) of evoked brain responses. In the figure, highly-correlated electroencephalographic (EEG) auditory event-related potentials (ERPs) **(b)** averaging responses to stimulus onsets recorded from three scalp channels **(a)** are separated by ICA **(c)** into the independent activities of three ICA brain sources **(d)**. The ICA training network **(c)** consists of a single layer of weights **W** summed through a sigmoidal nonlinearity **g()**. The rows of the trained ICA weight matrix **W** are linear spatial filters each passing activity arising from a single ICA source. Filtering the input data with these filters produces the event-related activity waveforms for each source **(d)**. Columns of the inverse weight matrix represent the projections of each source on the sensor array, here indicated by interpolated scalp potential maps **(e)**.

Figure 2 - ICA decomposition of averaged 14-channel ERPs obtained from one subject in a sustained auditory detection experiment[4].

(a) Responses to *(blue traces)* detected (n=209) and *(red traces)* undetected (n=81) slow-onset noise-burst targets.

(b) Source activity waveforms for the resulting 14 ICA sources during detected (*blue traces*) and undetected (*red traces*) response epochs. Seven sources (here labeled D1-3 and U1-4) are activated for 50-300 ms during one or both responses, while three sources (S1-3) compose the auditory steady-state response (SSR)[22] to a click train presented throughout the experiment at one eighth the EEG sampling rate. The four unlabeled sources may be mixtures of multiple weak event-related and spontaneous EEG and noise sources.

(c) Projected activity of sources U1-4 *(colored traces)* superimposed on the scalp waveforms of the undetected-target response (from a) *(grey traces).* Note that source U1 accounts for the central positive peak near 300 ms, whereas the succeeding negativity is decomposed into three source components (U2-4).

(d) *(grey traces)* Scalp waveforms from the detected-target response (a) and *(colored traces)* projected scalp activity of three ICA sources (D1-3).

METHODS: The subject pressed a button each time he heard a noise-burst target (mean frequency, 10/min; duration, 350 ms; rise time, 150 ms; intensity, 6 dB sensation level) embedded in a continuous stream of 62 dB white noise during a 28-min session. Brief, task-irrelevant tones (50 ms, 72 dB, 568 Hz) also were presented at 2-4 s intervals. EEG was collected (sampling rate, 312.5 Hz; pass band 0.1-100 Hz) from thirteen scalp electrodes referred to the right mastoid, and from a bipolar diagonal electroocular (EOG) placement. See [4] for further details. ICA input was the two 1-sec (312-point) ERP vectors shown in (a). with block size (presented data points per learning step) 10 and learning rate annealed from 0.0006 to 0.00001 during training.

Figure 3 - Projected scalp topographies and reliability of the labeled ICA sources in Fig. 2b.

(a) *(heads)* Scalp topographies of the major ICA sources (D1-3) of the detected-target response. *(traces)* Source waveforms for sources D1-3 in two sessions from the same subject, separately analyzed (and colored as in Fig. 2d). Source D1 has a left central distribution, while source D2 is active at the median response time (406 ms) and has midline parietal and precentral extrema compatible with an origin in or near motor cortex. Source D3 accounts for portion of the late EOG activity in the ERP

record.

(b) *(heads)* Projected scalp topographies and replicability of ICA sources U1-4 of the undetected-target response. *(traces)* Source waveforms for sources U1-4 in the same two sessions, analyzed separately and colored as in Fig. 2c. Sources U1-2 have similar central distributions, while sources U3-4 have prefrontal or extraocular extrema.

(c) ICA decomposition of the 39-Hz auditory SSR in the detected-target ERP (Fig. 2a). *(heads)* Topographic maps, individually scaled. *(leftmost traces)* SSR scalp topography at its maximum (16 ms). *(right traces)* Projected scalp topographies for ICA sources S1-3. S1 has a bilateral frontal distribution compatible with bilateral sources in the auditory cortices[23], while S2 has a bilateral parietal distribution. S3 is largest in EOG and prefrontal channels. Together, sources S1-3 account for 95.3% of total variance in the input SSR.

METHODS: SSR waveforms were derived from the detected-target ERP (Fig. 2a) by averaging successive 25.6-ms ERP time segments. Topographic mapping used spherical spline interpolation[24].

Figure 4 - Stability of the ICA source decomposition.

(a) Reproducibility of ICA source waveforms obtained from separate analyses of detected-target *(blue)* and undetected-target *(red)* ERPs from the session shown in Fig. 2a and from a second session for the same subject. To test the dependence of the ICA decomposition on the sensor array used in the analysis, different subsets of 11 of the 14 scalp channels were used in each analysis *(head plots)*. Highly similar source waveforms appeared in both analyses.

(b) ICA source waveforms and topographies of ICA source D2 (Fig. 2b) analogues obtained in five separate decompositions of *(left column)* the grand-mean detected-target response from 11 subjects and *(right columns)*responses from two separate sessions on each of 2 subjects.

(c) Stability of the source waveform for source D1 recovered in separate analyses of input data sets consisting of from one to three 1-s ERPs *(see key). ICA decompositions of each input data set* included one source whose activity and scalp topography (not shown) in response to detected targets matched that of source D1.

(d) Effect of reducing the number of scalp sensors on the stability of source U1. As the number of randomly-selected EEG channels in the input data was decreased from 14 to 4, an ICA source was repeatedly obtained whose activity waveform *(solid red line)* and projected scalp activity *(dashed blue line)* in the selected channels were correlated with those of source U1. Median correlations for 100 randomly-selected channel subsets are shown. In the same analyses, most subsets of 6 or more input channels produced sources strongly resembling sources U2, U3, D1, D2, and S1.

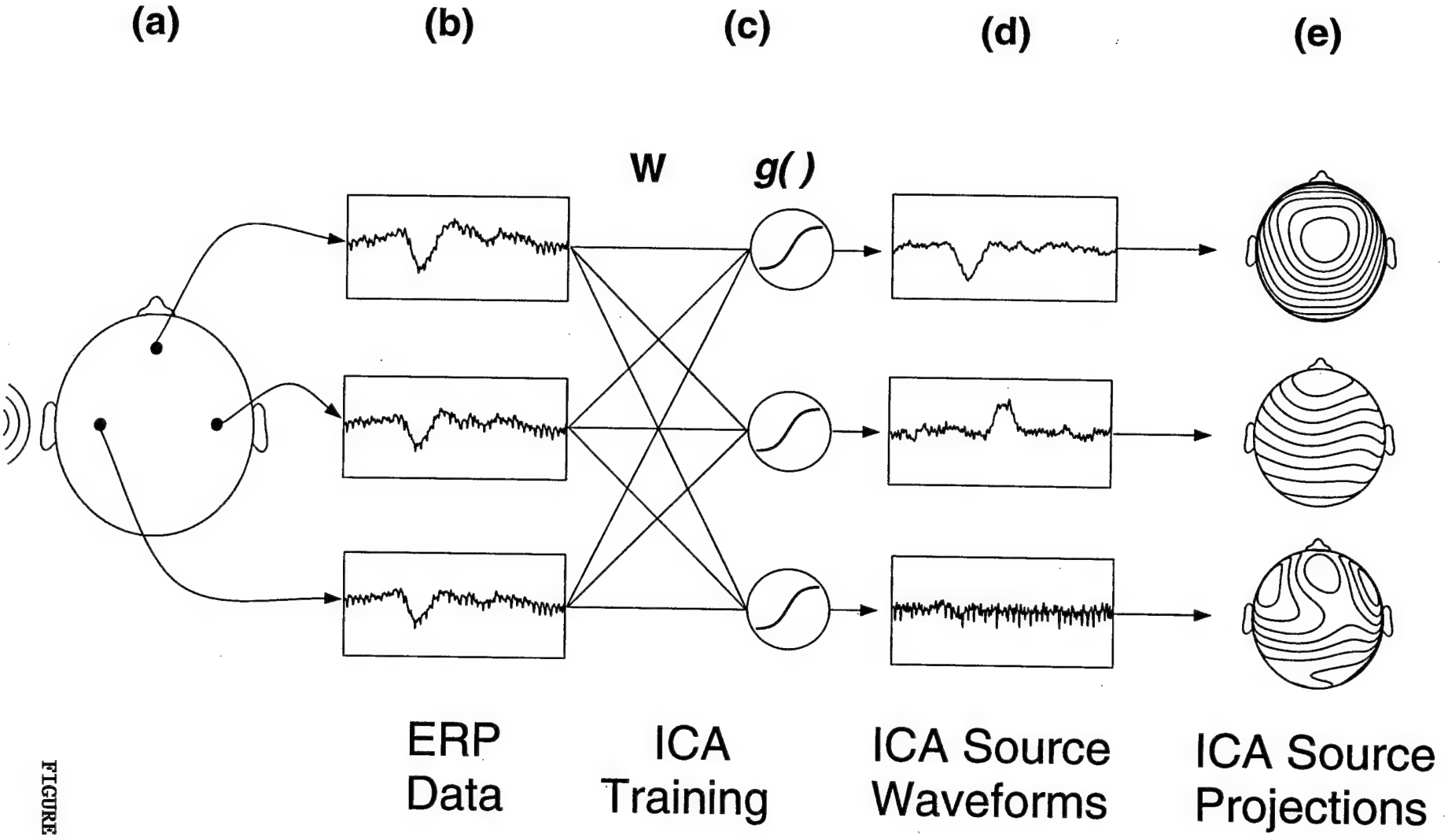
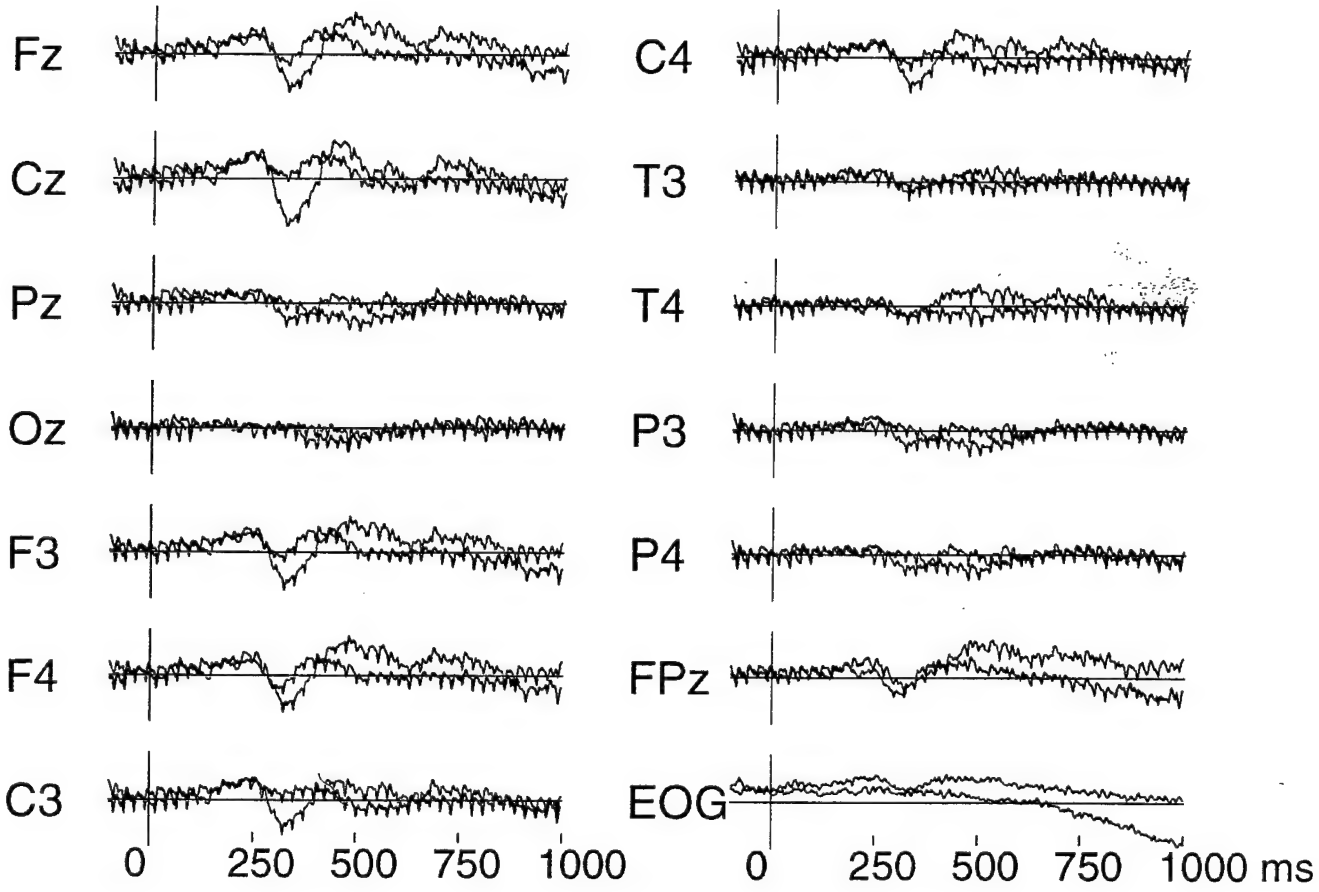
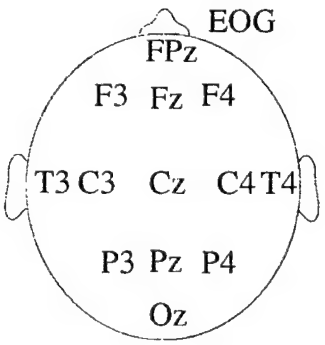

(a)
(b)
(c)
(d)
(e)
W
g( )
ERP Data
ICA Training
ICA Source Waveforms
ICA Source Projections

FIGURE 1

Scalp ERPs





-12 uV

+12 uV

FIGURE 2(a)

ICA Sources

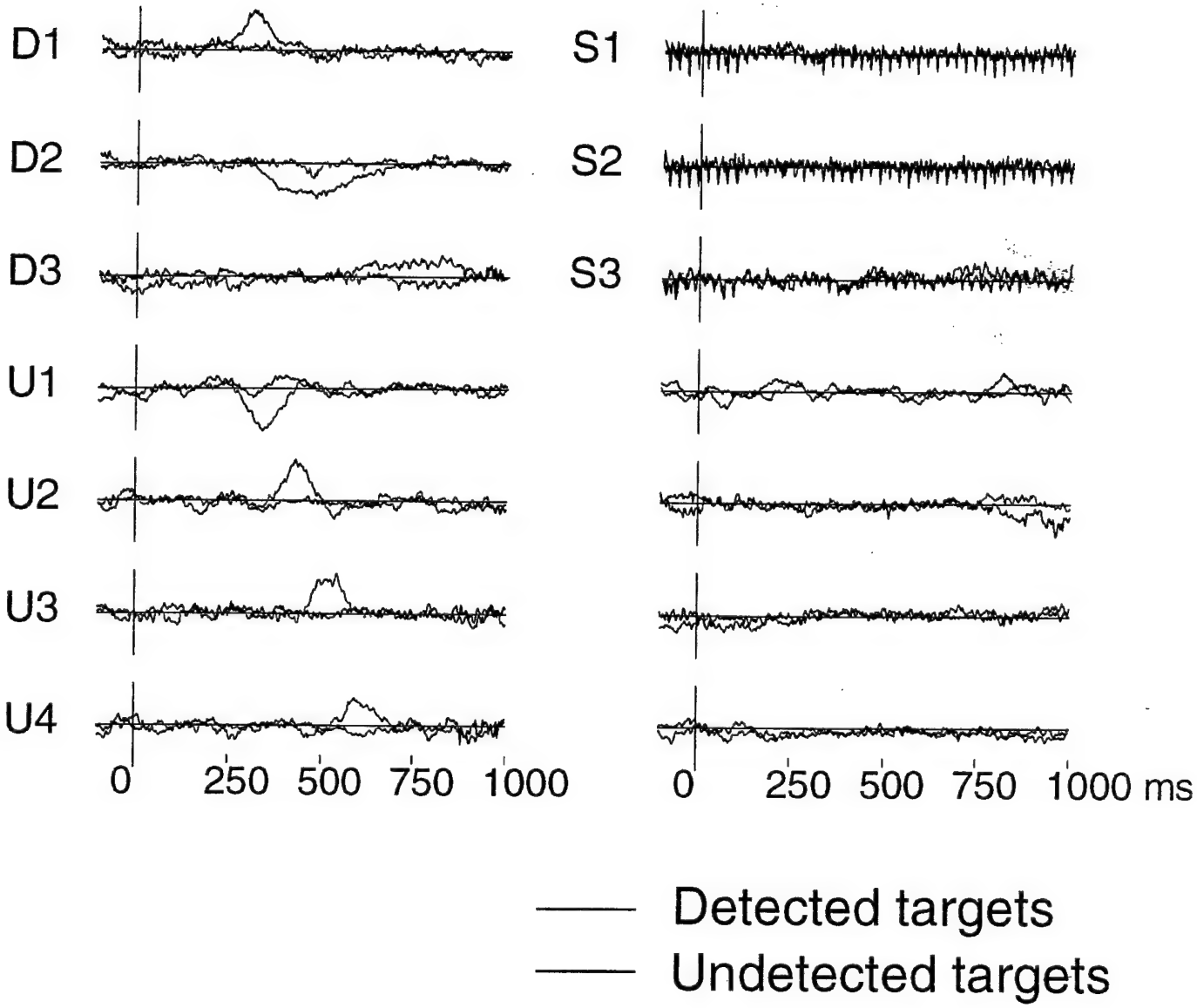


FIGURE 2(b)

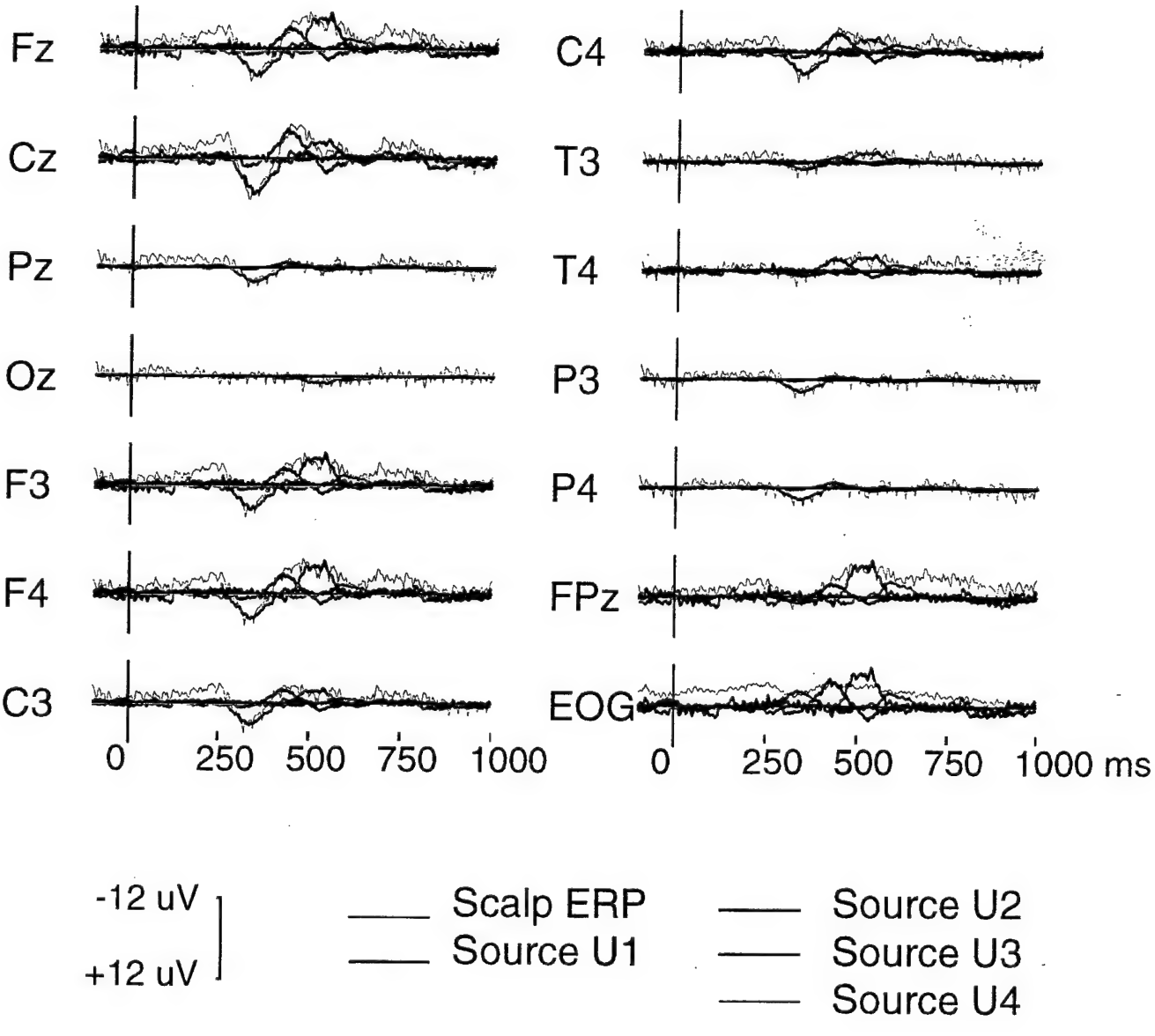

Source Projections (Undetected Targets)
Fz
Cz
Pz
Oz
F3
F4
C3
C4
T3
T4
P3
P4
FPz
EOG
0
250
500
750
1000
0
250
500
750
1000 ms
-12 uV
+12 uV
Scalp ERP
Source U1
Source U2
Source U3
Source U4

FIGURE 2(c)

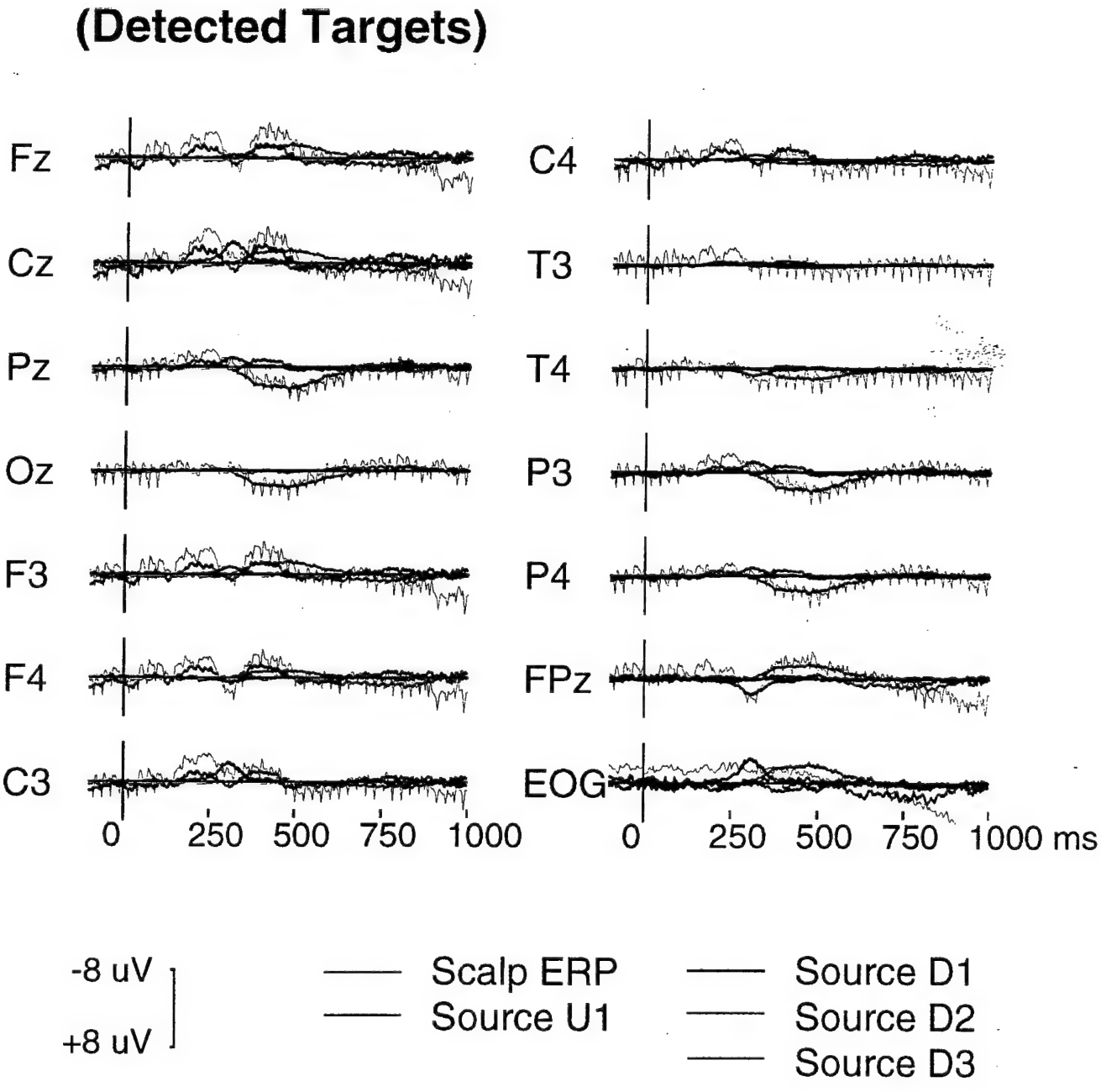

(Detected Targets)
Fz
Cz
Pz
Oz
F3
F4
C3
C4
T3
T4
P3
P4
FPz
EOG
0
250
500
750
1000
0
250
500
750
1000 ms
-8 uV
+8 uV
Scalp ERP
Source U1
Source D1
Source D2
Source D3

FIGURE 2(d)

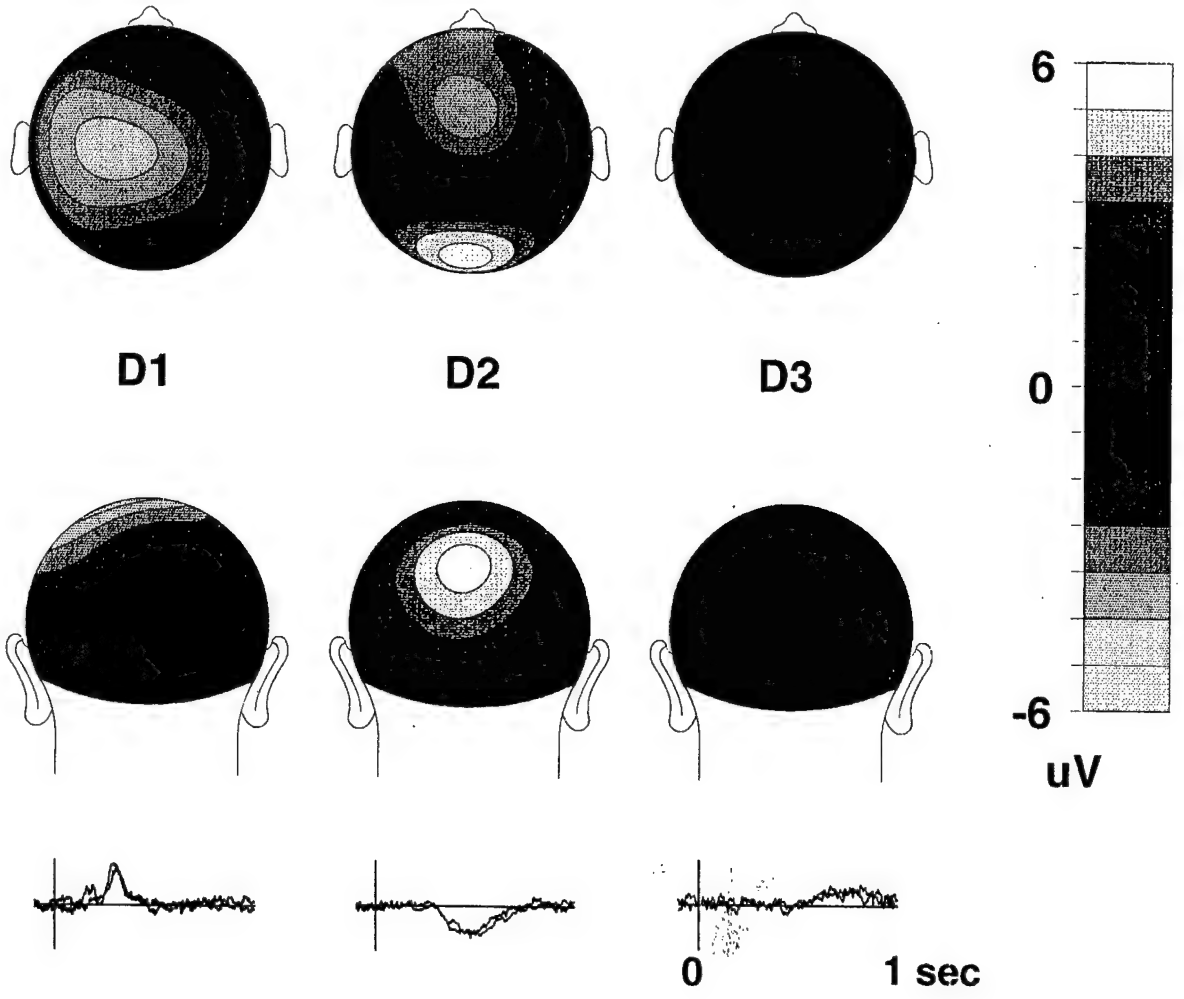


FIGURE 3(a)

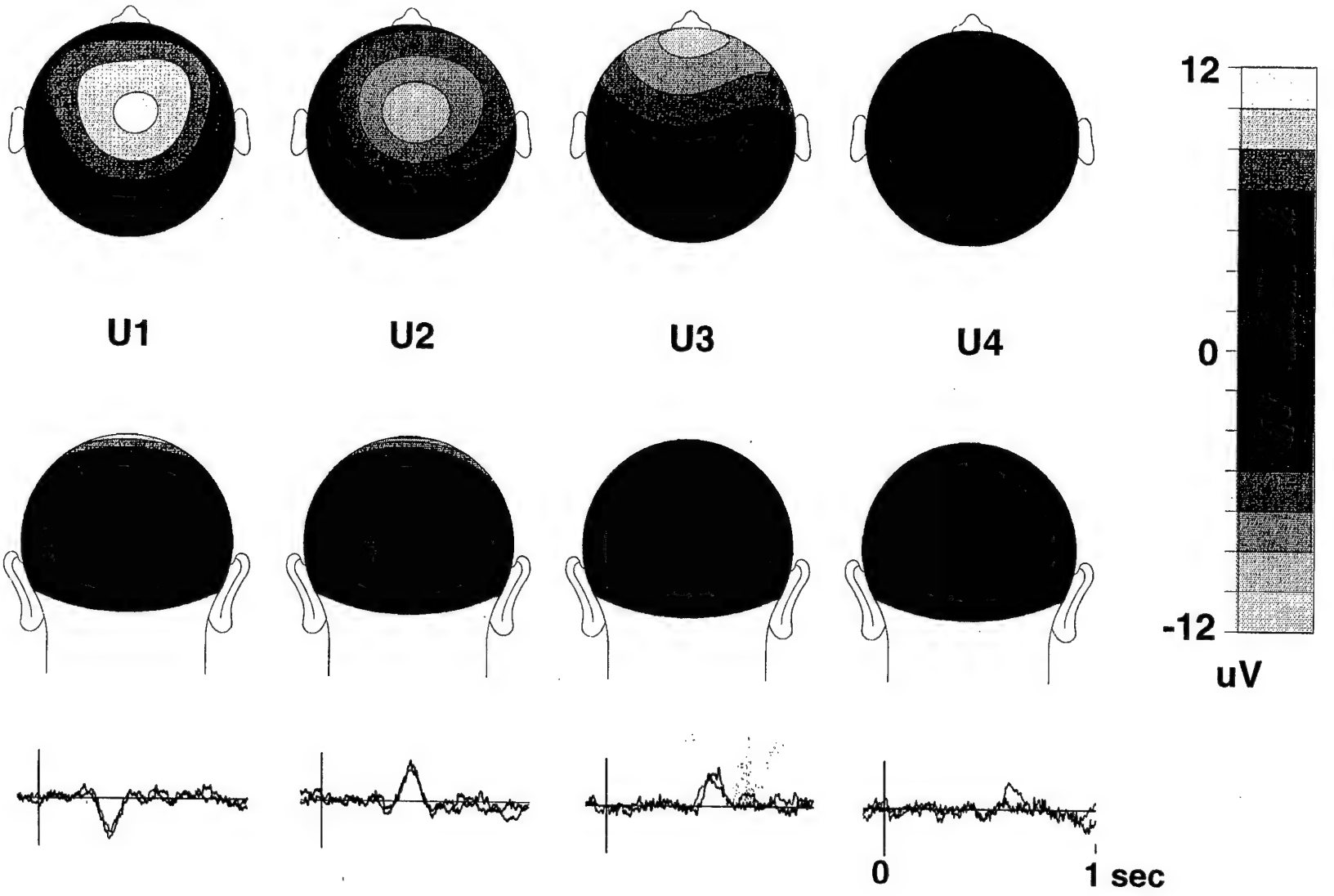


FIGURE 3(b)

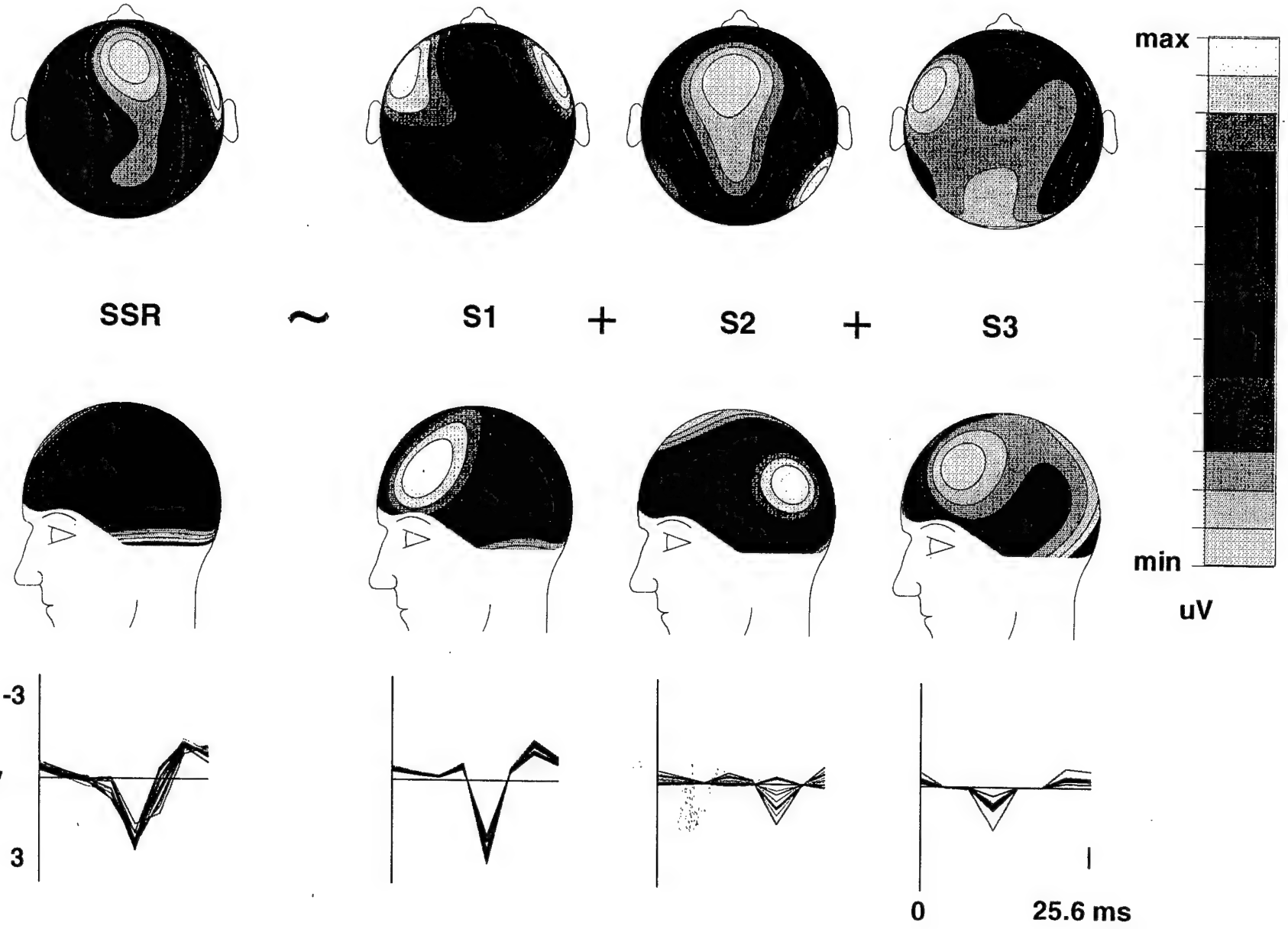


FIGURE 3(c)

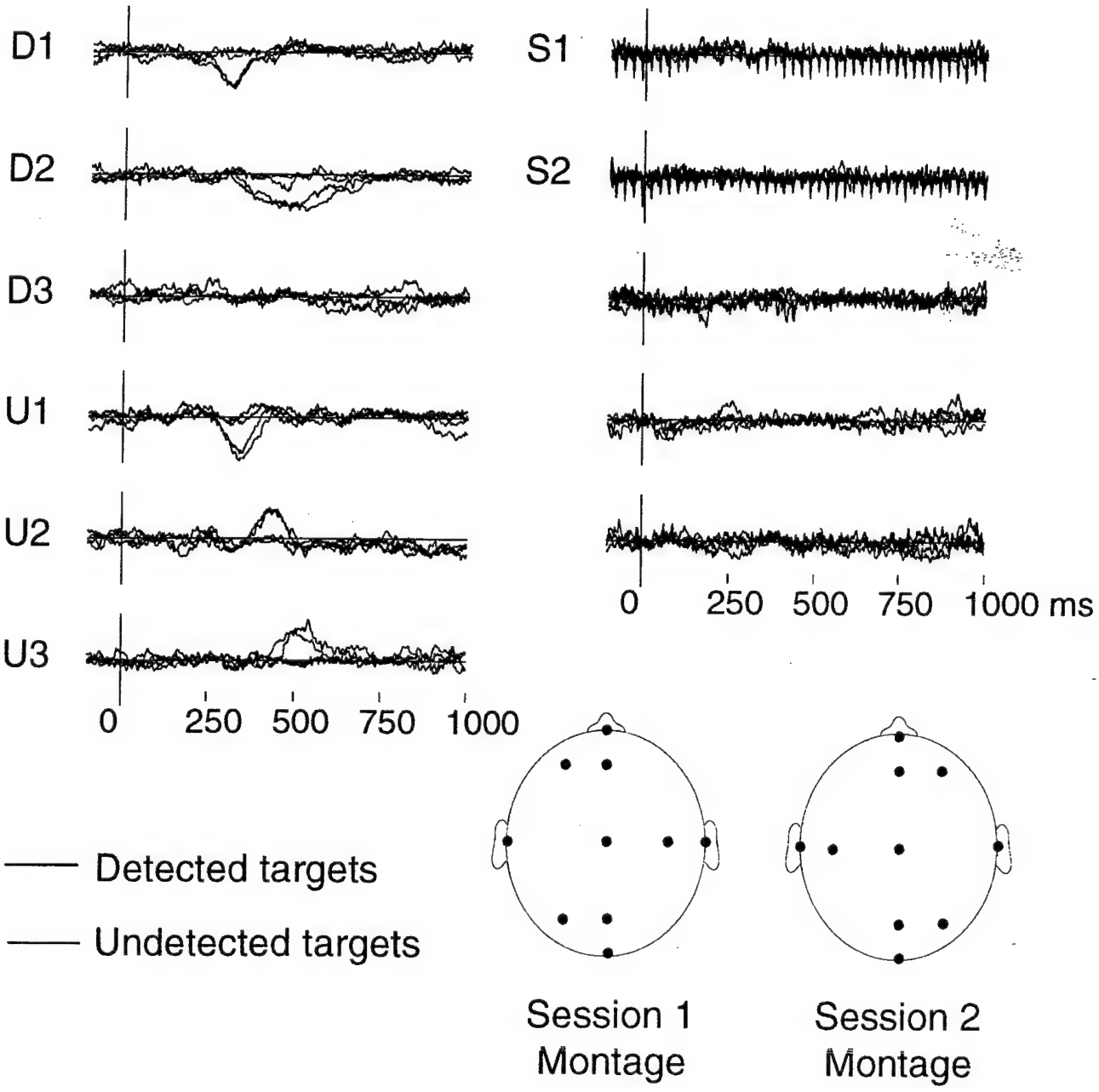

Sensors
D1
D2
D3
U1
U2
U3
S1
S2
0
250
500
750
1000
1000 ms
Detected targets
Undetected targets
Session 1
Montage
Session 2
Montage

FIGURE 4(a)

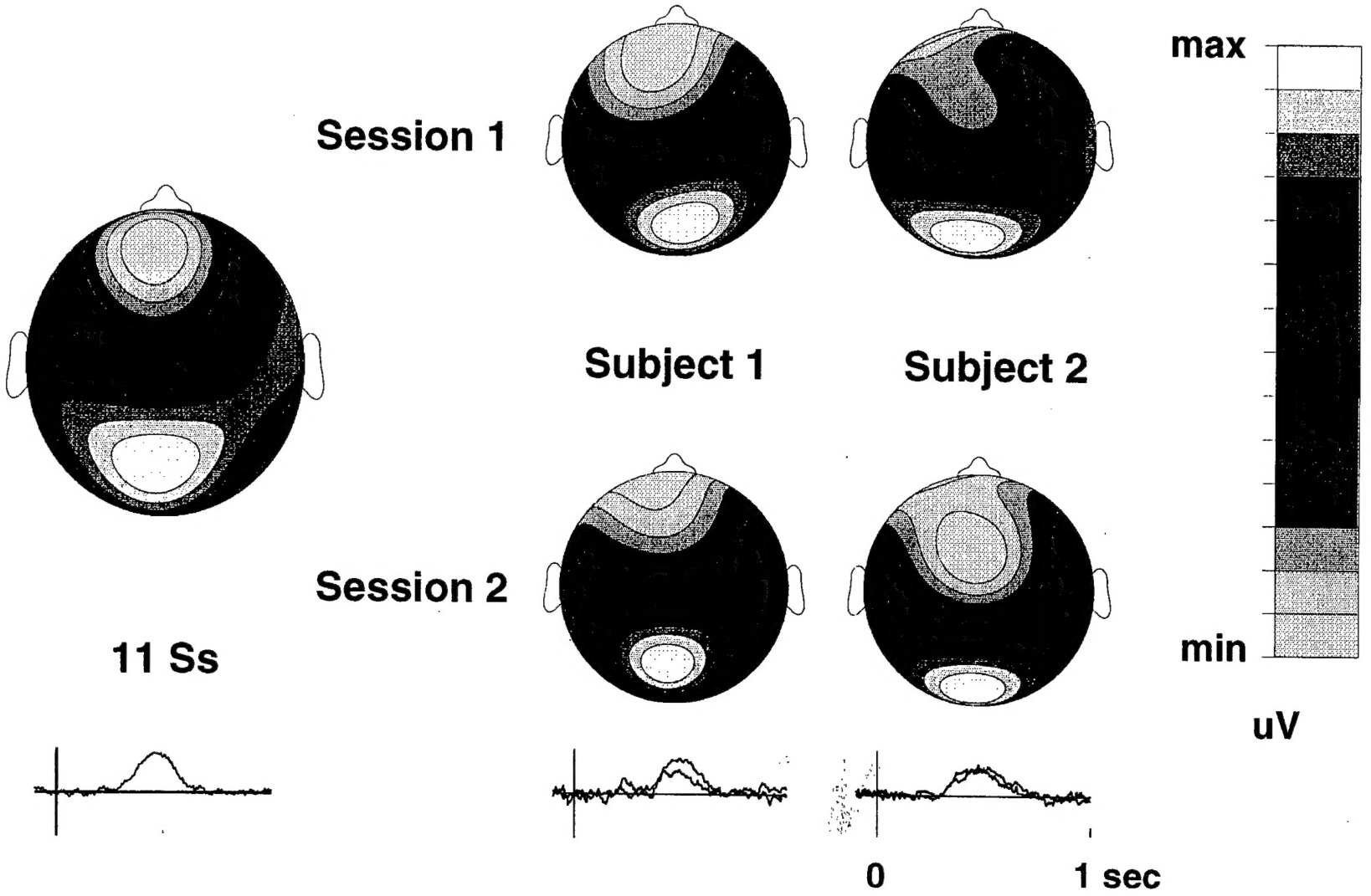

Subjects and Sessions
Source D2
Session 1
Session 2
Subject 1
Subject 2
11 Ss
max
min
uV
0
1 sec

Figure 4(b)

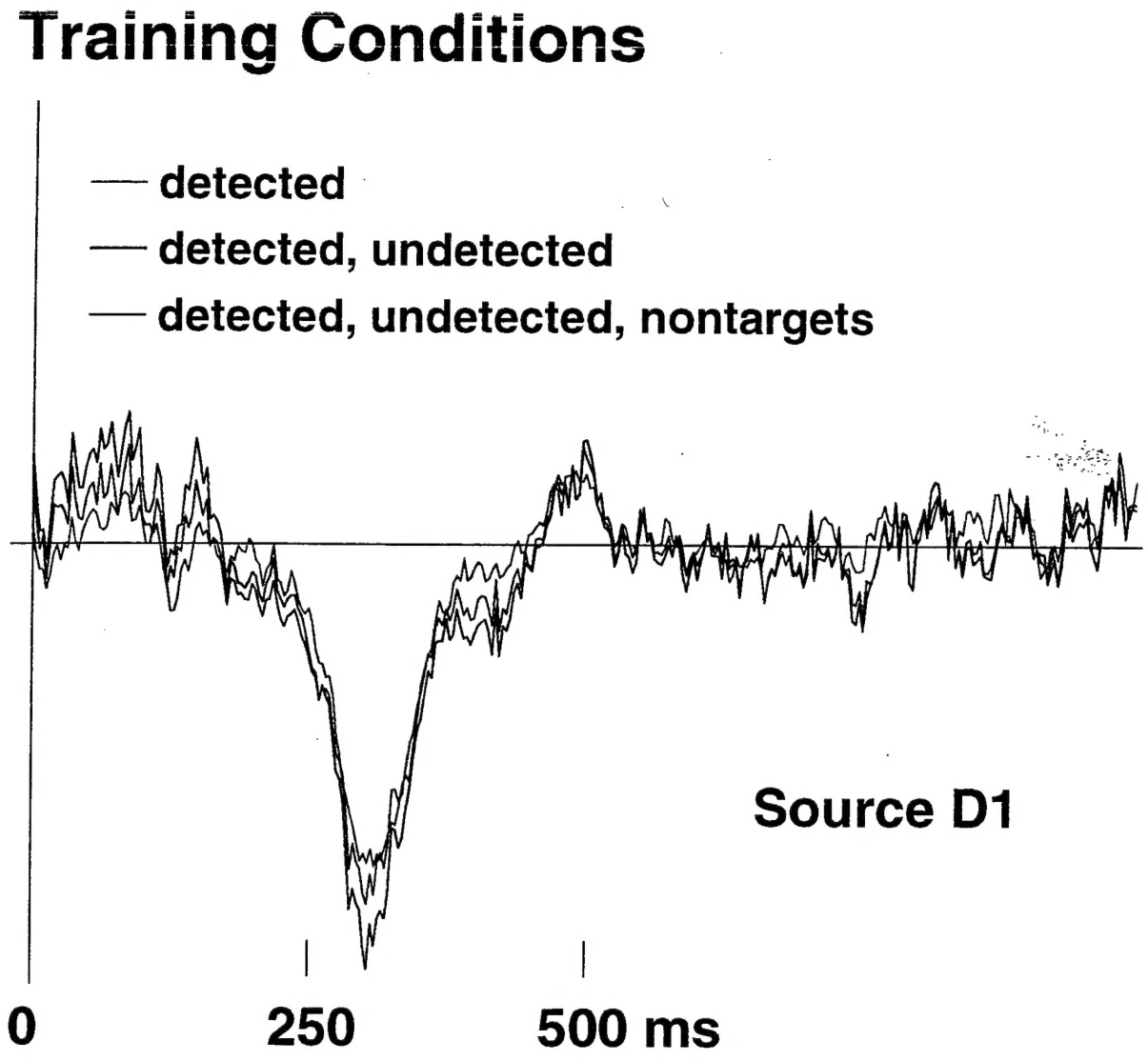

Training Conditions
detected
detected, undetected
detected, undetected, nontargets
Source D1
0
250
500 ms

FIGURE 4(c)

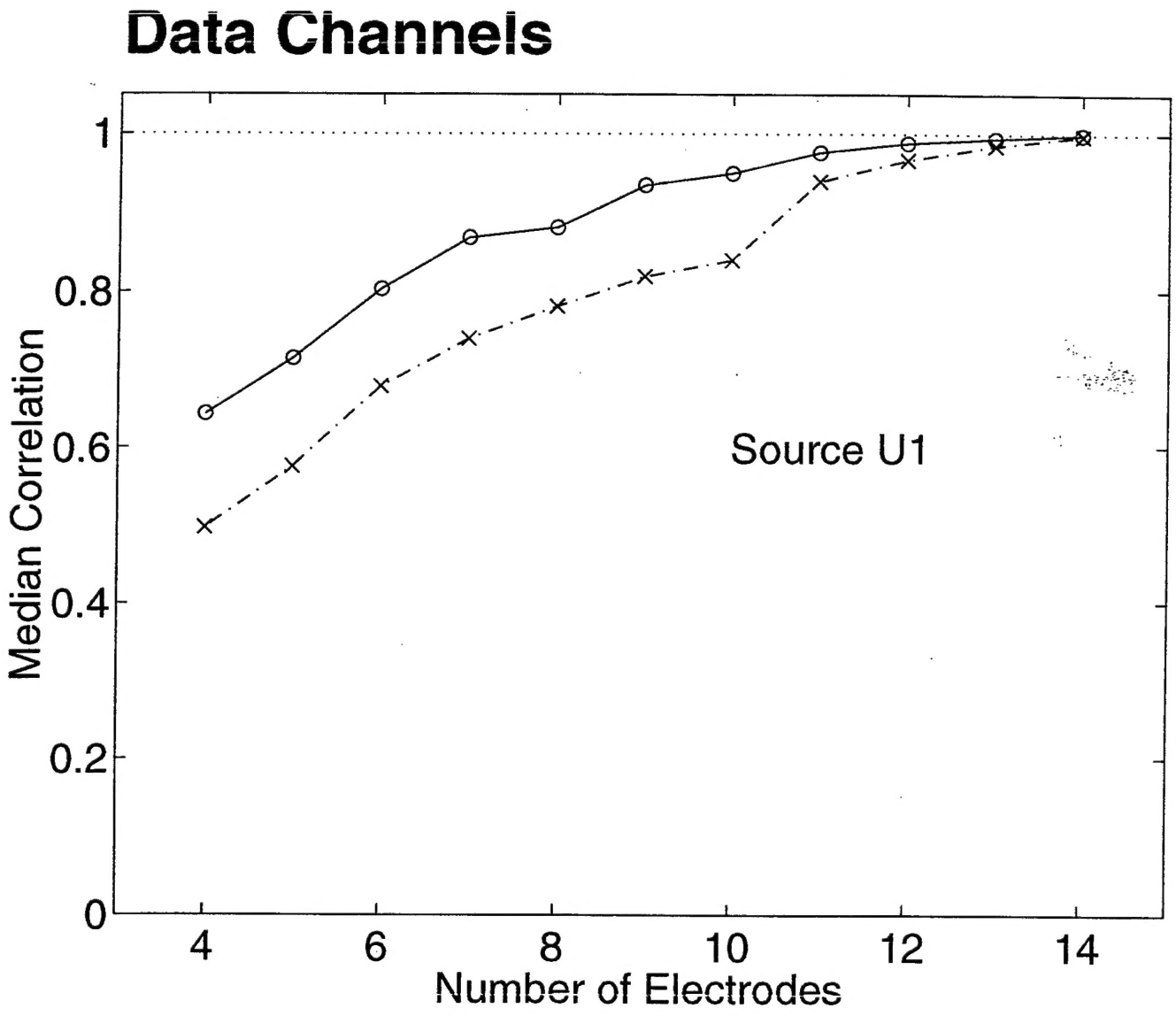

Data Channels
Median Correlation
1
0.8
0.6
0.4
0.2
0
4
6
8
10
12
14
Number of Electrodes
Source U1

FIGURE 4(d)

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE & DATE OVERED
	June 1996	Final

4. TITLE AND SUBTITLE
Blind Separation of Event-related Brain Responses into Independent Components

5. FUNDING NUMBERS
Program Element:
ONR Reimbursable
Work Unit Number: 9610

6. AUTHOR(S)
Scott Makeig, Tzyy-Ping Jung, Anthony J Bell, Dara Ghahremani, Terrance J. Sejnowski

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Health Research Center
P.O. Box 85122
San Diego, CA 92186-5122

8. PERFORMING ORGANIZATION
Report 96-10

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)
Naval Medical Research & Development Command
National Naval Medical Center NCR
Bethesda, MD 20814-5044

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
1997 Proceedings, National Academy of Science (in press)

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE
A

13. ABSTRACT ***(Maximum 200 words)***

Functional imaging of brain activity based on changes in blood flow does not supply information about the relative timing of brief bursts of neural activity in different brain areas. Multichannel electric or magnetic recordings from the scalp provide high temporal resolution, but are not easily decomposed into the separate activities of multiple brain networks. We report here a method for the blind separation of event-related brain responses into spatially stationary and temporally independent subcomponents using an Independent Component Analysis algorithm. Applied to electroencephalographic responses from an auditory detection task, each of the most active identified sources accounted for all or part of a previously identified response component. This spatiotemporal decomposition was robust to changes in sensors and input data length, and was stable within subjects. The method can be used to assess the timing, strength, and stability of event-related activity in brain networks during cognitive tasks, regardless of source location.

14. SUBJECT TERMS
Event-related brain responses

15. NUMBER OF PAGES
25

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	**Unclassified**	**Unclassified**	**Unclassified**